Drinking water quality, what can we learn from data?

G. Kyritsakas, V. Speight, C. Thom and J. Boxall

Stream

The Industrial Doctorate Centre for the Water Sector

Background

- Water Companies are required to have a monitoring sample program throughout their Water Distribution Systems (WDS) (Fig.1) to measure chemical and biological parameters and guarantee safe drinking water to their customers.
- Samples results are stored creating datasets that increase yearly. Scottish Water (SW) is taking more than 300000 samples per year from their assets.
- These datasets are not easily accessible and time is required to collect, clean and prepare the data for use.
- Machine Learning (ML) is a Computer Science domain that creates algorithms to allow computers to learn from the data and examples and make estimations of future behaviours or reveal hidden correlations between various parameters.

The aim of this project is to investigate ,with the use of ML techniques, what useful information the water quality (WQ) datasets contain regarding WQ parameters and characteristics and their correlation with water deterioration in WDS.

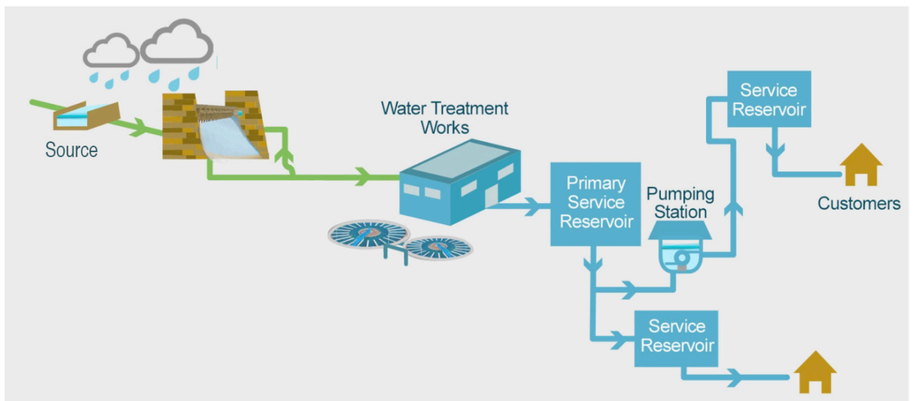

Figure 1 An example of a Water distribution System

Objectives

- To apply ML techniques and investigate the factors of water quality deterioration in WDS and predict future water behavior.
- To identify which ML techniques are appropriate for either predicting future water quality behavior or understanding the roots of WQ deterioration.
- To suggest ways that will facilitate the daily application of ML techniques inside SW, creating a supporting tools that assist pro-active decisions for interventions.

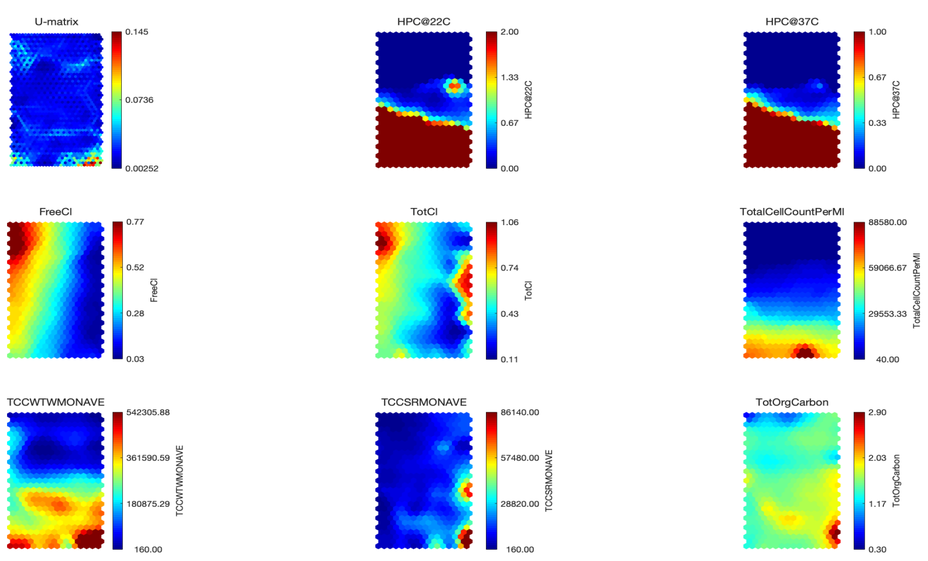

Figure 2 **An example of a ML technique**

Self Organizing Maps (SOMs)[1] is a clustering and correlations technique. In my research, SOMs were applied for the identification of the factors of bacteriological failures in the WDS.

Methodology

Step 1
Gathering and collecting water quality data

Step 2
Cleaning datasets – removing unwanted data

Step 3
Selecting part of the dataset for a specific WQ problem

Step 4
Selecting the appropriate ML technique (depending on the output requirement)

Step 5
Evaluating results and ML performance

Step 6a
Predicting behavior and identifying future risks

Step 6b
Finding parameters' correlations and factors of deteriorations

MACHINE LEARNING PROCESS

Benefits

- Provide to SW customers clean and safe drinking water
- Predict future WQ deterioration events
- Support decision makers in their pro-active decisions for interventions
- Reduce customer complains and increase SW reputation

Reference

Kohonen, T. 1990 The self-organizing map. Proceedings of the IEEE 78, 1464–1480.

www.stream-idc.net

The University Of Sheffield.

Scottish Water
Trusted to serve Scotland

EPSRC
Engineering and Physical Sciences Research Council

For further information: gkyritsakas1@Sheffield.ac.uk
Postal Address: Sir Frederick Mappin Building |Mappin Street|S13JD Sheffield